FIGURE 1 - General Overview of Distributed File Storage System
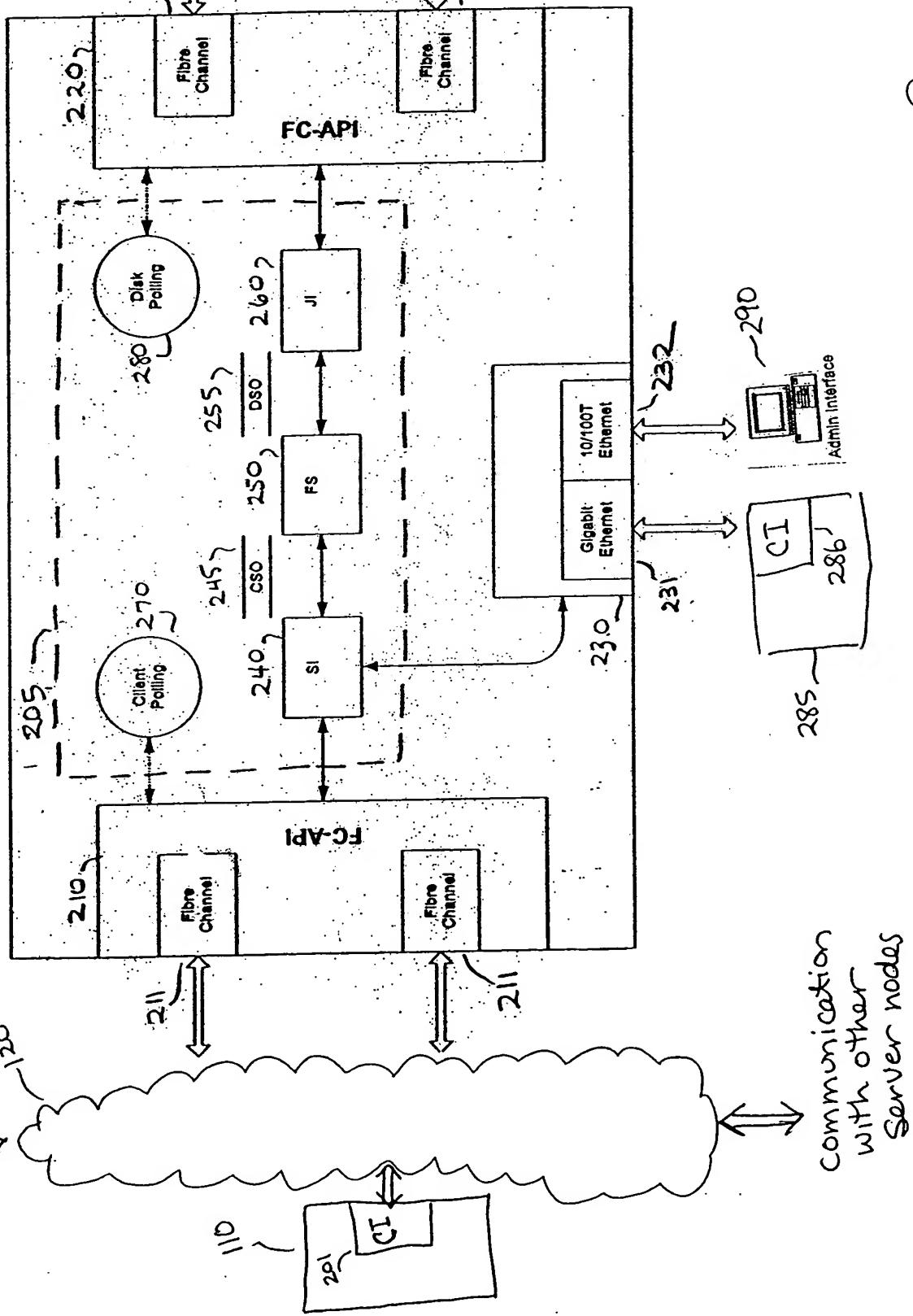
FIGURE 2 : One Embodiment of a Server Node

G-node Table ~ 33[0]

contains information about File attributes

(See Figure 6)

Filename Table 310

contains information about File names

(See Figure 4)

Gee Table 320

contains information about File block location

(See Figure 5)

Cache Node Table 350

contains information about cached file block

(See Figure 8)

Gnid Table 340

contains information about Directory Structure

(See Figure 7)

points to a G-node

points to a Gee-string

points to the file's Gnode and to the Parent's G-node

points to a Child Gnid Index

points to a filename entry

Points to a Gee

points to a Cache Node

Other Servers' G-node Tables (for child files stored on other servers)

FIGURE 3 - Five metadata structures
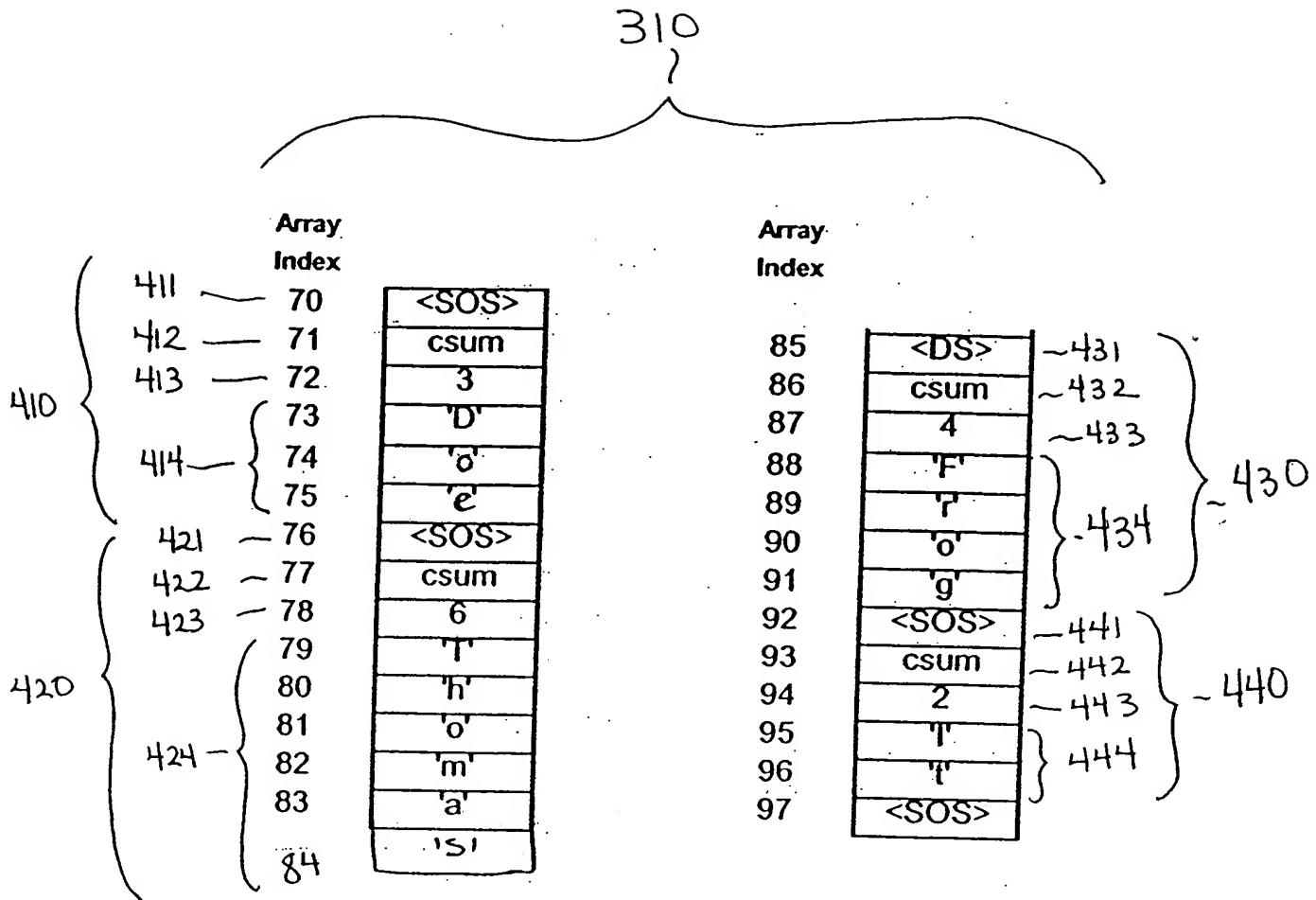
FIGURE 4 - Sample Portion of a Filename Table

320

	Index	G-Code	Data	File Logical Block
510	45	GNODE	Gnode = 67, Extent = 2, Root = TRUE	
511	46	DATA	Disk Logical Blocks: 456, 457 Drive 13	1
512	47	DATA	Disk Logical Blocks: 667, 668 Drive 15	2
513	48	DATA	Disk Logical Blocks: 112, 113 Drive 19	3
514	49	PARITY	Disk Logical Blocks: 554, 555 Drive 2	
515	50	DATA	Disk Logical Blocks: 458, 459 Drive 13	4
516	51	DATA	Disk Logical Blocks: 669, 670 Drive 15	5
517	52	DATA	Disk Logical Blocks: 119, 120 Drive 19	6
518	53	PARITY	Disk Logical Blocks: 556, 557 Drive 2	
519	54	LINK	Index 76	
	...	...	...	
520	76	GNODE	Gnode = 67, Extent = 3, Root = FALSE	
521	77	DATA	Disk Logical Blocks: 460, 461, 462 Drive 13	7
522	78	DATA	Disk Logical Blocks: 671, 672, 673 Drive 15	8
523	79	PARITY	Disk Logical Blocks: 121, 122, 123 Drive 19	
524	80	LINK	Index 88	
	...	...	...	
525	88	GNODE	Gnode = 67, Extent = 3, Root = FALSE	
526	89	DATA	Disk Logical Blocks: 463, 464, 465 Drive 13	9
527	90	DATA	Disk Logical Blocks: 674, 675, 676 Drive 15	10
528	91	PARITY	Disk Logical Blocks: 124, 125, 126 Drive 19	
529	92	GNODE	Gnode = 43, Extent = 4, Root = FALSE	
	...	...	...	

590 591 592

550 551 552 500

FIGURE 5 · Sample Portion of a Gee Table

Attribute Data
602 — File Attribute – type
604 — File Attribute – mode
606 — File Attribute – links
608 — File Attribute – uid
610 — File Attribute – gid
612 — File Attribute – size
614 — File Attribute – used
620 — File Attribute – fileId
622 — File Attribute – atime
624 — File Attribute – mtime
626 — File Attribute – ctime
628 — Child Gnid Index
630 — Gee Index – Last Used
631 — Gee Offset – Last Used
632 — Gee Index – Midpoint
633 — Gee Offset – Midpoint
634 — Gee Index – Tail
635 — Gee Offset – Tail
636 — Gee Index – Root
638 — Gnode Status
640 — Quick Shot Status
642 — Quick Shot Link

600

FIGURE 6 - G-NODE ATTRIBUTES

FIGURE 7 - Structure of a Gnid String

DataGee	
810 —	
815 — PrevPtr	
820 — NextPtr	
825 — CacheBlockAddr	
ReadCt	
830 —	
835 — CacheTime	
840 — Regenerated	
845 — CacheBlockHiAddr	
850 — CacheBlockLoAddr	

800
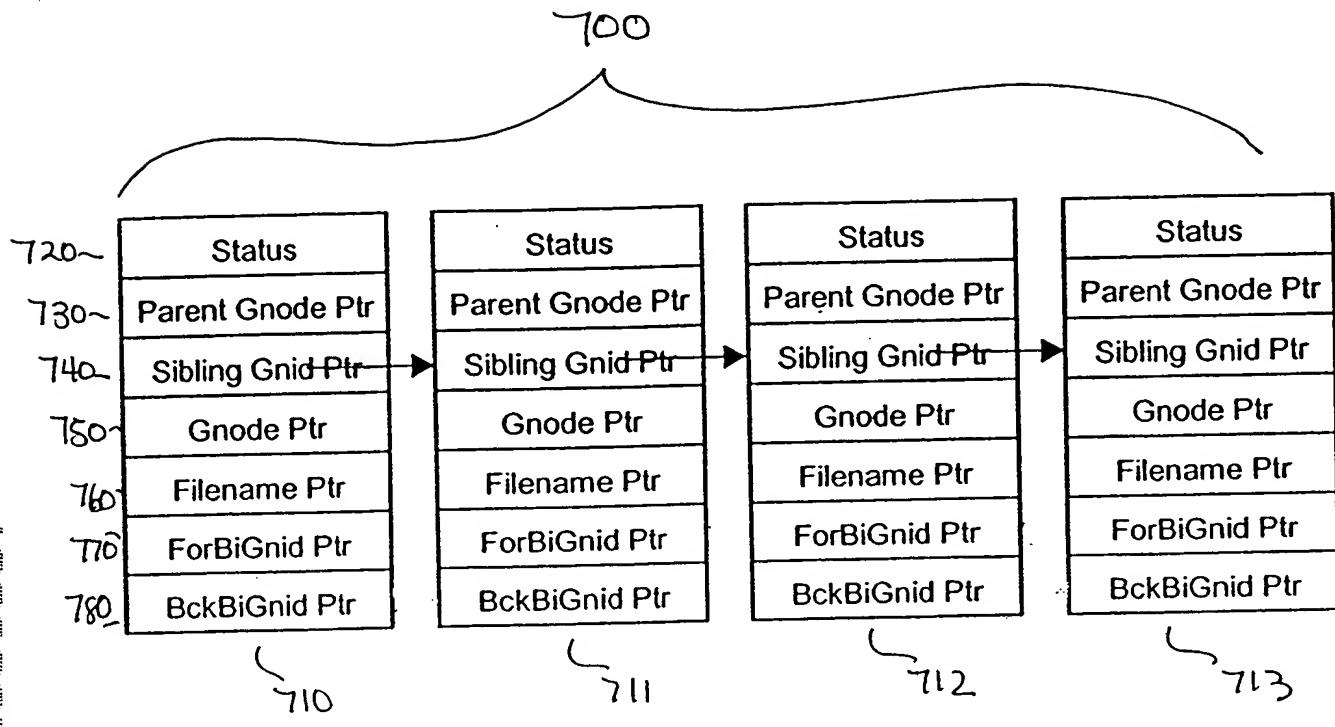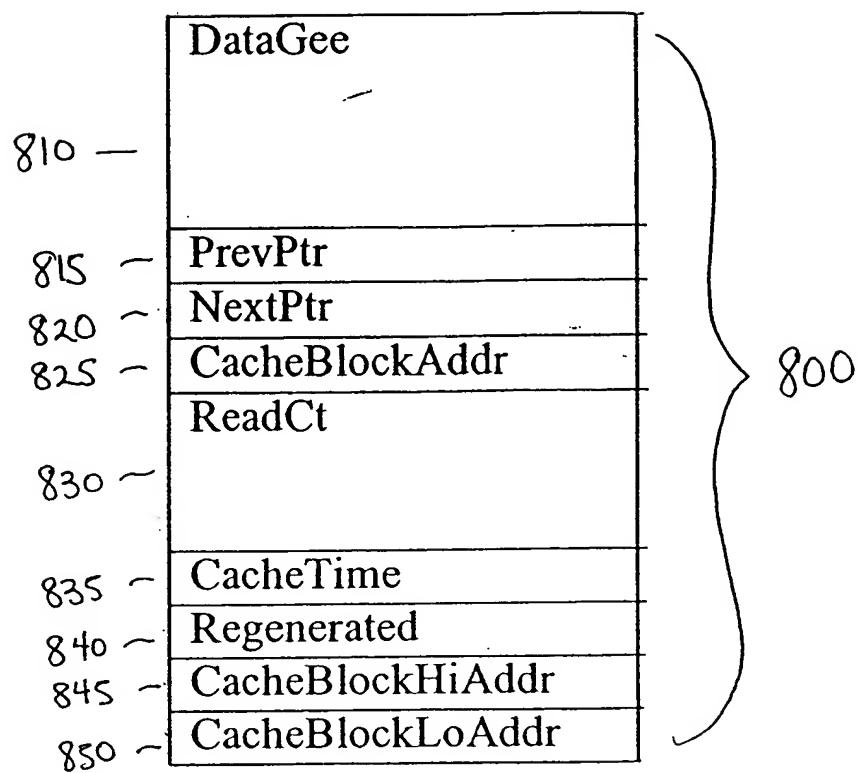
FIGURE 8a - Structure of a Cache Node

350

Normal List

Write List Cache Nodes are returned to "top" of Normal List

Most recently used

875

Read List cache nodes are returned to "top" of Normal List

Write List

Normal List Cache Nodes may come from anywhere on Normal List

Read List

Normal List Cache Nodes may come from anywhere on Normal List

880

Least recently used

870

865

860
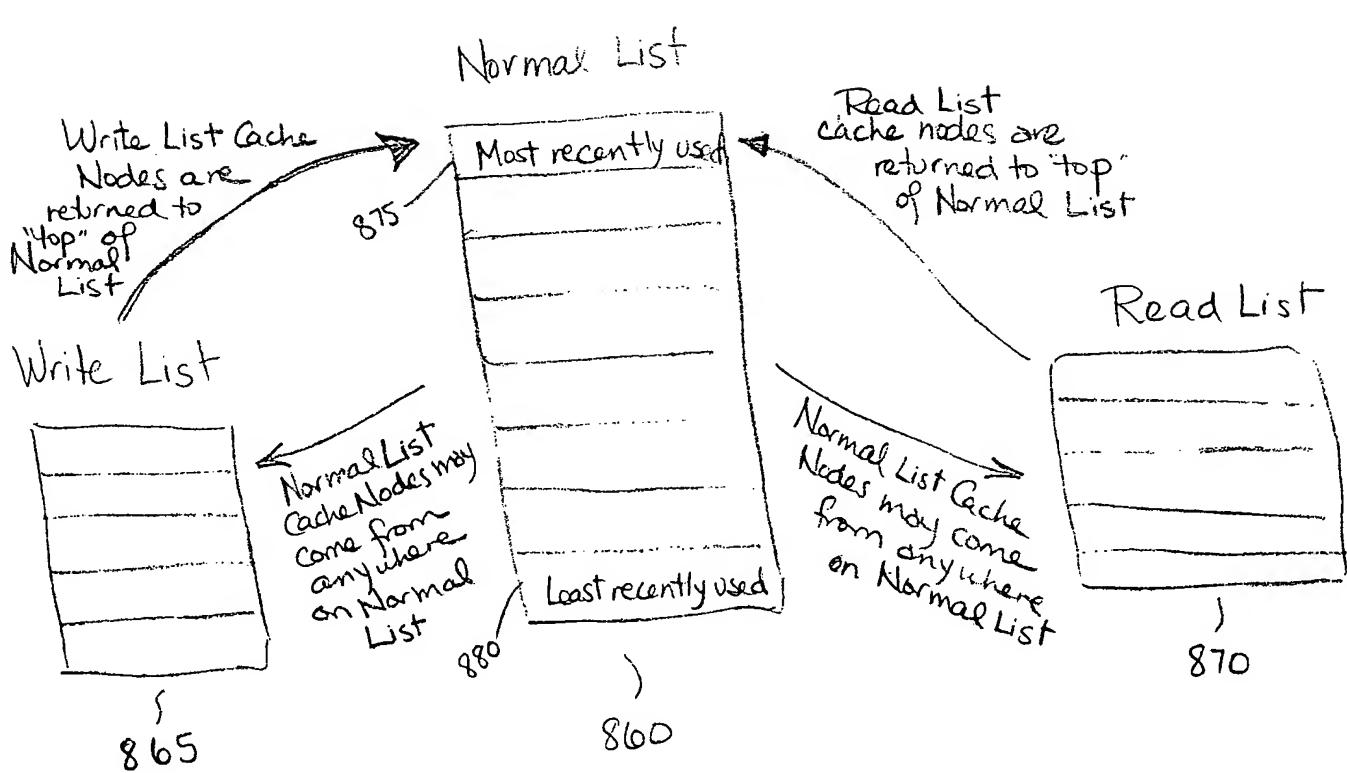
FIGURE 8B – Conceptual division of a Cache Node Table into Three Lists

900

911   912   921   922   923

930 931 932 933 934 935 936 937	lockStatus timeoutCt next pending mode access clientId uid	lockStatus timeoutCt next pending mode access clientId uid	lockStatus timeoutCt next pending offset length exclusive clientId uid svid	lockStatus timeoutCt next pending offset length exclusive clientId uid svid	lockStatus timeoutCt next pending offset length exclusive clientId uid svid

938 offset
939 length
940 exclusive
941 clientId
942 uid
943 svid

910

920

FIGURE 9 - A Sample Lock String

1020

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1010

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1030

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1000

1001 clientId
1002 counter
1003 ltLink
1004 gtLink
stackNext
stackPrev

1040

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1050

clientId
counter
ltLink
gtLink
stackNext
stackPrev

FIGURE 10 · Refresh Nodes configured as a binary tree.

1100

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1110

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1120

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1130

clientId
counter
ltLink
gtLink
stackNext
stackPrev

1105 stackNext
1106 stackPrev
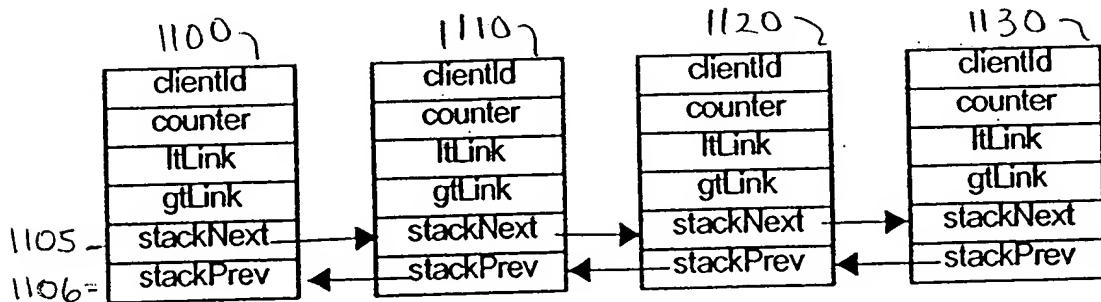
FIGURE 11 - RefreshNodes configured as a
doubly-linked list

~1200

Status	
1210 ~	Status
1220 ~	Intent Type
1230 ~	goalBufferIndex
1240 ~	spare
1250 ~	driveSector
1260 ~	drive
1270 ~	intentData
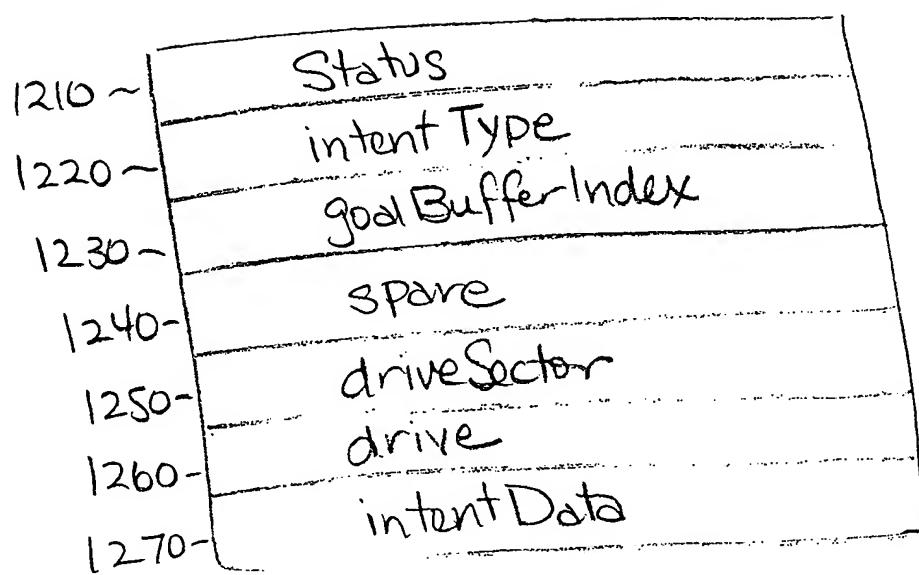
FIGURE 12 - Structure of an Intent Log Entry

1300

1310	1320	1330
Recommended NIC	ServerID of the file's primary owner	Gnode Index

BIT 31   BIT 29   BIT 23   BIT Ø
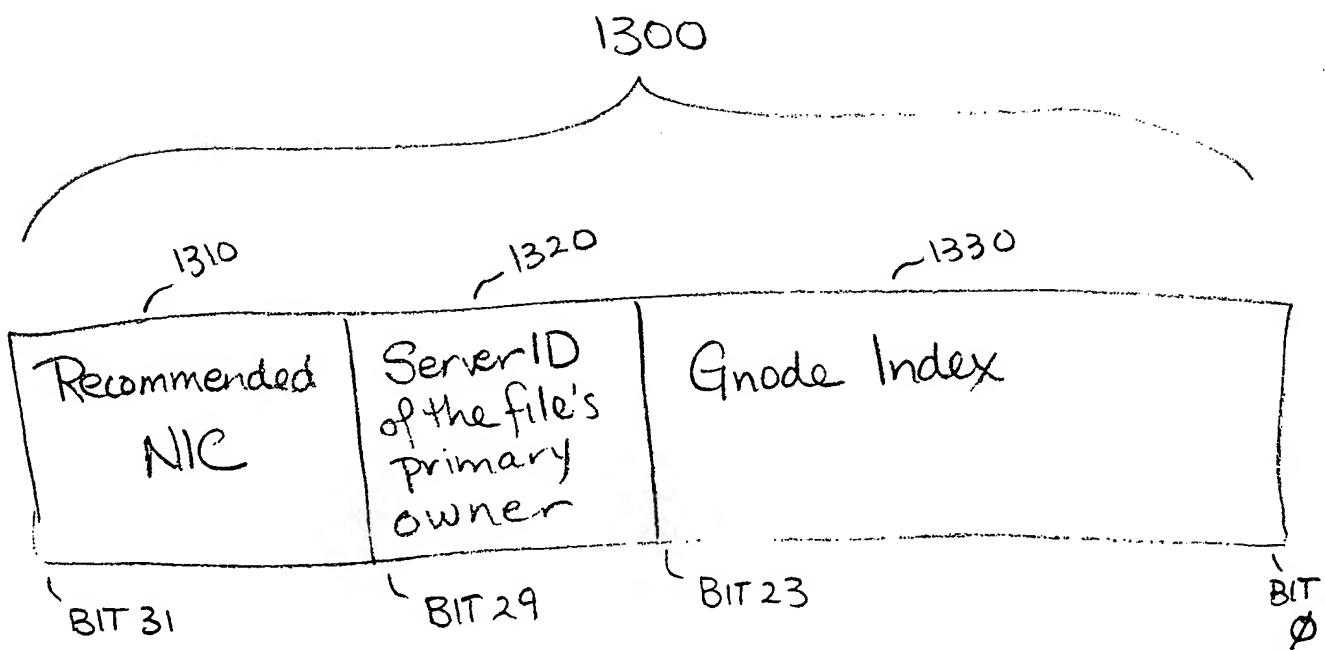
FIGURE 13 - Structure of a File Handle

FIGURE 14a: Example of a File Look-Up

FIGURE 14b Example of a File Access

START 1500

1500

Does Client have desired file's file handle ? — 1510

No → Client and Server Perform File Handle Look-up (See Figure 16) 1515

YES

Client sends FileAccess Request to Server Indicated in File Handle — 1520

Server Accesses Gnode Indicated in File Handle — 1525

Server Accesses Appropriate Gee Indicated in Gnode — 1530

Is G-code CACHE PARITY or CACHE DATA ? — 1535

No → Server Caches File Data from Disk (See Figure 17) 1540

YES — 1545

Server Accesses Cache Node Indicated in Gee — 1545

Server Manipulates Cache Nodes as Needed — 1550

Server Accesses File Data in Cache Location Indicated in Cache Node — 1555

Server Performs Desired Operation on File Data — 1560

END — 1570

FIGURE 15:
Performing a File Access

START ~1605

Client sends Look-Up Request to Server ~1610

Server Accesses Gnode Indicated in Look-Up Request ~1615

Server Uses Child Gnid Index in Gnode to Access the Directory's Gnid-String ~1620

Server Calculates Checksum and Filename Length for Desired next Pathname Component ~1625

Server Uses Filename Ptr field in Gnid to access Filename Entry ~1630

Is checksum from Filename Entry greater than checksum from Look-Up Request ? ~1635

YES → Return error message to Client: File Not Found ~1640

Do checksums and filename lengths match ? ~1645

YES → Do filenames match ? ~1650

YES → Return File Handle to Client ~1655

END. File Handle Look-Up ~1665

NO → Access next Gnid on Gnid-String ~1660

NO

1600

FIGURE 16: Performing a File Handle Look-Up

1705 ~ (START)

1540
)

1710 ~ Identify Least Recently Used Cache Node

1720 ~ Write associated file data from cache to disk

1730 ~ Return Data Gee from Cache Node to Gee Table

1740 ~ Copy Data Gee for new data into Cache Node

1750 ~ Copy new data from disk to Cache

1760 ~ (END)

FIGURE 17: Caching File Data

START —1805

1810 — Client sends file allocation request including parent directory's file handle and filename of desired file

1815 — Server indicated in parent's file handle receives request

1820 — Parent server uses workload statistics from other server nodes to decide: Will "parent" server own the new file ?

NO

1850 — "Parent" Server sends request for file allocation to a "second" server

1855 — "Second" server allocates new Gnode and Filename Entry and creates a new file

1860 — "Second" server sends file handle for new file back to "parent" server

YES

1830 — "Parent" Server allocates new Gnode and creates new file

1835 — "Parent" server creates new Gnid and Filename Entry

1840 — "Parent" Server sends file handle for new file to requesting client

1845 — END

1800

FIGURE 18 — File Allocation

- Gnode Redirectors (GNR)

1900

gnr[n]
...
gnr[11]
gnr[10]
gnr[9]
gnr[8]
gnr[7]
gnr[6]
gnr[5]
gnr[4]
gnr[3]
gnr[2]
gnr[1]
gnr[0]

1910

Not In Cache
Not In Cache
Not In Cache
Not In Cache
Not In Cache
Not In Cache

File Handle is this Index

Not In Cache

Not In Cahce
Not In Cache

1920

cached gnode - C
...
cached gnode - 2
cached gnode - 1
cached gnode - 0

600

FIGURE 19

2000

2010 — STATUS

2020 — LINKING INFORMATION

GNODE

2030 —

2040 — File location data

128 Bytes

16 KBytes

Figure 20a

FIGURE 20b

CONVENTIONAL RAID MAPPING
(PRIOR ART)

FIGURE 21

FIGURE 22A

110

100

CLIENT   CLIENT   CLIENT   CLIENT

COMMUNICATION FABRIC
120

SERVER
130

150

DISK ARRAY
140

FIGURE 22B

CLUSTER

SERVER   SERVER

DISK ARRAY

DISK ARRAY

160

SERVER   SERVER

DISK ARRAY

DISK ARRAY

160

SERVER   SERVER

DISK ARRAY

DISK ARRAY

160

110

100

CLIENT   CLIENT   CLIENT   CLIENT

COMMUNICATION FABRIC
120

130

SERVER   SERVER
131

150

151

165

DISK ARRAY
140

DISK ARRAY
141

DISK ARRAY

160

FIGURE 23

FILE #1 — 2320

| DATA$_1$ 1 | DATA$_2$ 1 | | | DATA$_5$ 1 | DATA$_6$ 1 | DATA$_7$ 1 |

2330

DATA$_3$ 1   DATA$_4$ 1

FILE #2 — 2340

| DATA$_1$ 2 | DATA$_2$ 2 | | |

2330

DATA$_3$ 2   DATA$_4$ 2

2335

| DATA$_1$ 1 | DATA$_2$ 1 | PARITY 1-2 1 | | | | | | DATA$_5$ 1 | DATA$_6$ 1 | DATA$_7$ 1 | PARITY 5-7 1 |

2330 2337   DATA$_3$ 1 PARITY3-4 1

DATA$_4$ 1

2335

| DATA$_1$ 2 | DATA$_2$ 2 | PARITY 1-2 2 | | | | |

2330 2337   DATA$_3$ 2 DATA$_4$ 2 PARITY 3-4 2

2305
DISK NUMBER

2300

2310

LOGICAL BLOCK OR CLUSTER

	0	1	2	3	4	5	6	7	8	9	10	11
0												
1				PARITY 1-2 2				DATA$_5$ 1				PARITY 5-7 1
2	DATA$_1$ 1	DATA$_2$ 1								PARITY 1-2 1		
3					DATA$_3$ 2							
4							PARITY 3-4 1					
5			DATA$_2$ 2									
6									DATA$_3$ 1		DATA$_7$ 1	
7				DATA$_4$ 1		DATA$_4$ 2						
8	DATA$_1$ 2									PARITY 3-4 2		
9			DATA$_6$ 1									
10												
11												

2337

2337

FIGURE 24A

```
2400

CLIENT STORAGE          2410
    REQUEST

        │ DATA    2415
        ▼

RECEIVE DATA FROM       2420
    CLIENT

        ▼

ANALYZE DATA /
DETERMINE BLOCK         2430
   DISTRIBUTION

        ▼

PARSE INTO DATA         2440
    BLOCK(S)

        ▼

DETERMINE PARITY
GROUP ARRANGEMENT       2450

        ▼

CALCULATE PARITY        2460
     BLOCKS

        ▼

DETERMINE BLOCK
DISTRIBUTION FOR        2470
     STORAGE

        ▼

 STORE BLOCKS           2480
```

FIGURE 24B

```
                    ┌─────────────────────┐
                    │   CLIENT STORAGE    │──── 2410
                    │      REQUEST        │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │    ANALYZE DATA /    │
                    │  DETERMINE BLOCK     │──── 2430
                    │    DISTRIBUTION      │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │   PARSE INTO DATA    │──── 2440
                    │      BLOCK(S)        │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │   DETERMINE PARITY   │──── 2450
                    │  GROUP ARRANGEMENT   │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │   DETERMINE BLOCK    │
                    │  DISTRIBUTION FOR    │──── 2470
                    │      STORAGE         │
                    └─────────────────────┘
                               │  DATA ──── 2415
                               ▼
                    ┌─────────────────────┐
                    │   RECEIVE DATA FROM  │──── 2420
                    │       CLIENT         │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │  CALCULATE PARITY    │──── 2460
                    │      BLOCKS          │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │    STORE BLOCKS      │──── 2480
                    └─────────────────────┘
```

2405

FIGURE 25

FIGURE 26A

PARITY GROUP A (BLOCK NUMBER= 4)

| DATA $_1$ 1 | → | DATA $_2$ 1 | → | DATA $_3$ 1 | → | DATA $_4$ 1 | - - - | PARITY1-4 1 |

2330 2335 2337 2502

PARITY GROUP B (BLOCK NUMBER= 2)

| DATA $_5$ 1 | → | DATA $_6$ 1 | - - - | PARITY 5-6 1 |

2330 2505 2337 2504

FIGURE 26B

PARITY GROUP C
(BLOCK NUMBER = 4 / BLOCK SIZE = 256K)

| DATA $_1$ 1 | → | DATA $_2$ 1 | → | DATA $_3$ 1 | → | DATA $_4$ 1 | - - - | PARITY1-4 1 |

2330 2335 2337 2506

PARITY GROUP C
(BLOCK NUMBER = 4 / BLOCK SIZE = 128K)

| DATA $_1$ 1 | → | DATA $_2$ 1 | → | DATA $_3$ 1 | → | DATA $_4$ 1 | - - - | PARITY1-4 1 |

2330 2505 2335 2337 2508

DISK ARRAY INITIALIZATION USING GEE TABLE
SPACE ALLOCATION

2530

INDEX	G-CODE	DATA
...	...	...
45	GNODE	EXTENT=2
46	DATA	BLOCKS 456, 457: Drive 13
47	DATA	BLOCKS 667, 668: Drive 15
48	DATA	BLOCKS 112, 113: Drive 19
49	PARITY	BLOCKS 554, 555: Drive 2
...	...	...
76	GNODE	EXTENT=3
77	DATA	BLOCKS 460, 461, 462: Drive 13
78	DATA	BLOCKS 671, 672, 673: Drive 15
79	PARITY	BLOCKS 121, 122, 123: Drive 19
...	...	...
88	GNODE	EXTENT=2
89	DATA	BLOCKS 463, 464, 465: Drive 2
90	DATA	BLOCKS 674, 675, 676: Drive 5
91	PARITY	BLOCKS 124, 125, 126: Drive 13
...		

2532   2534   2536   2542   2538   2540

FIGURE 27

ARRAY PREPARATION / G-TABLE FORMATTING

_~ 2448

2550~ DETERMINE DISK
CHARACTERISTICS

2555~ DETERMINE BLOCK / PARITY
GROUP ALLOTMENT

2560~ DETERMINE MAPPING / CREATE
GEE TABLE

2575~ DATA STORAGE / GEE TABLE
UPDATES
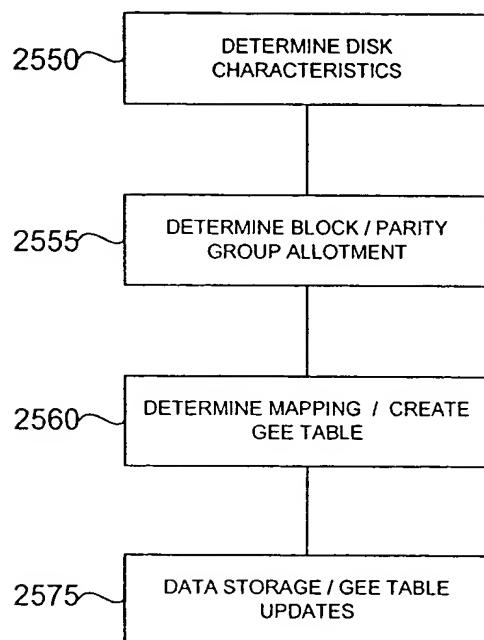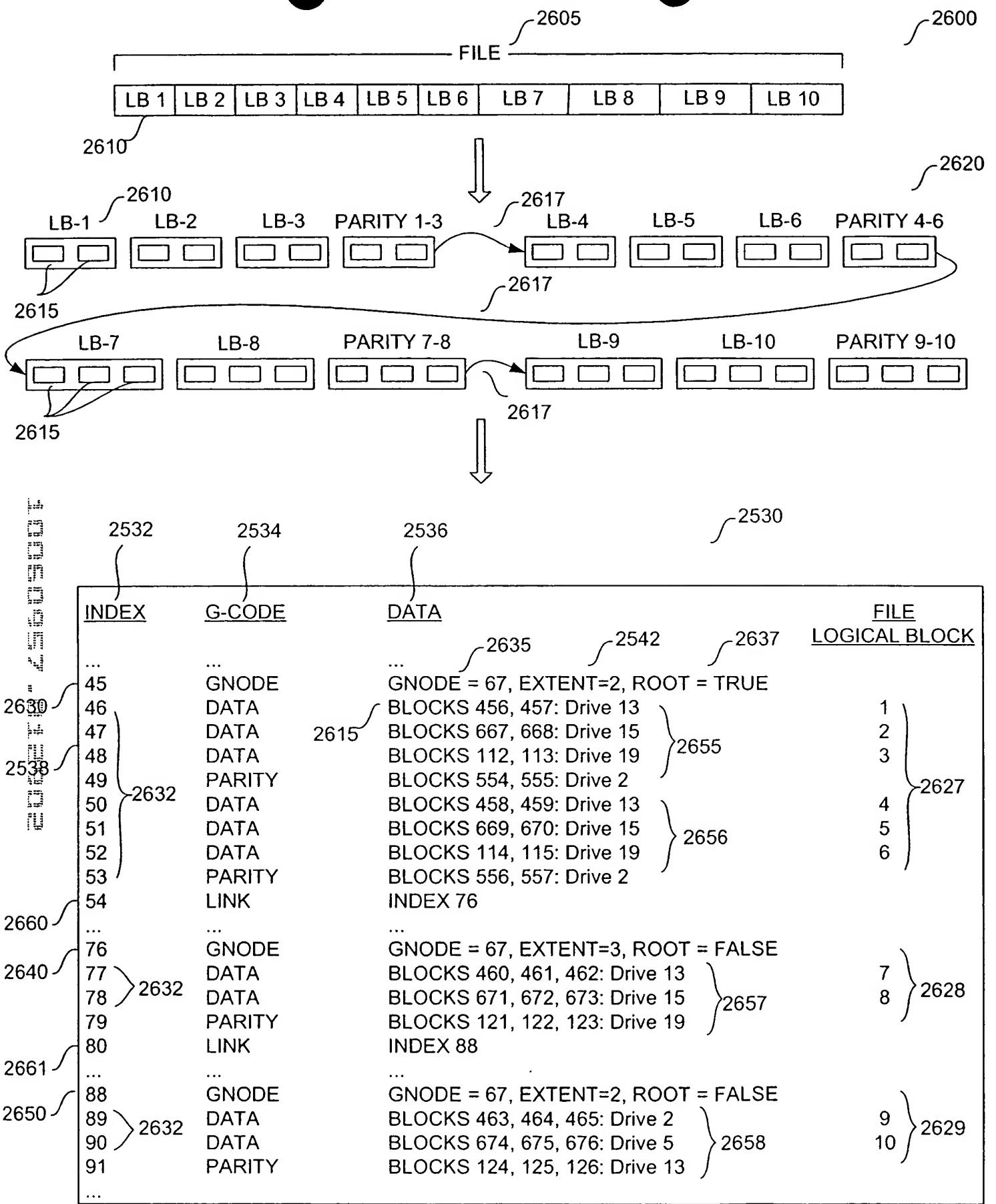
FIGURE 28

FIGURE 29

DRIVE FAILURE RECOVERY MECHANISM
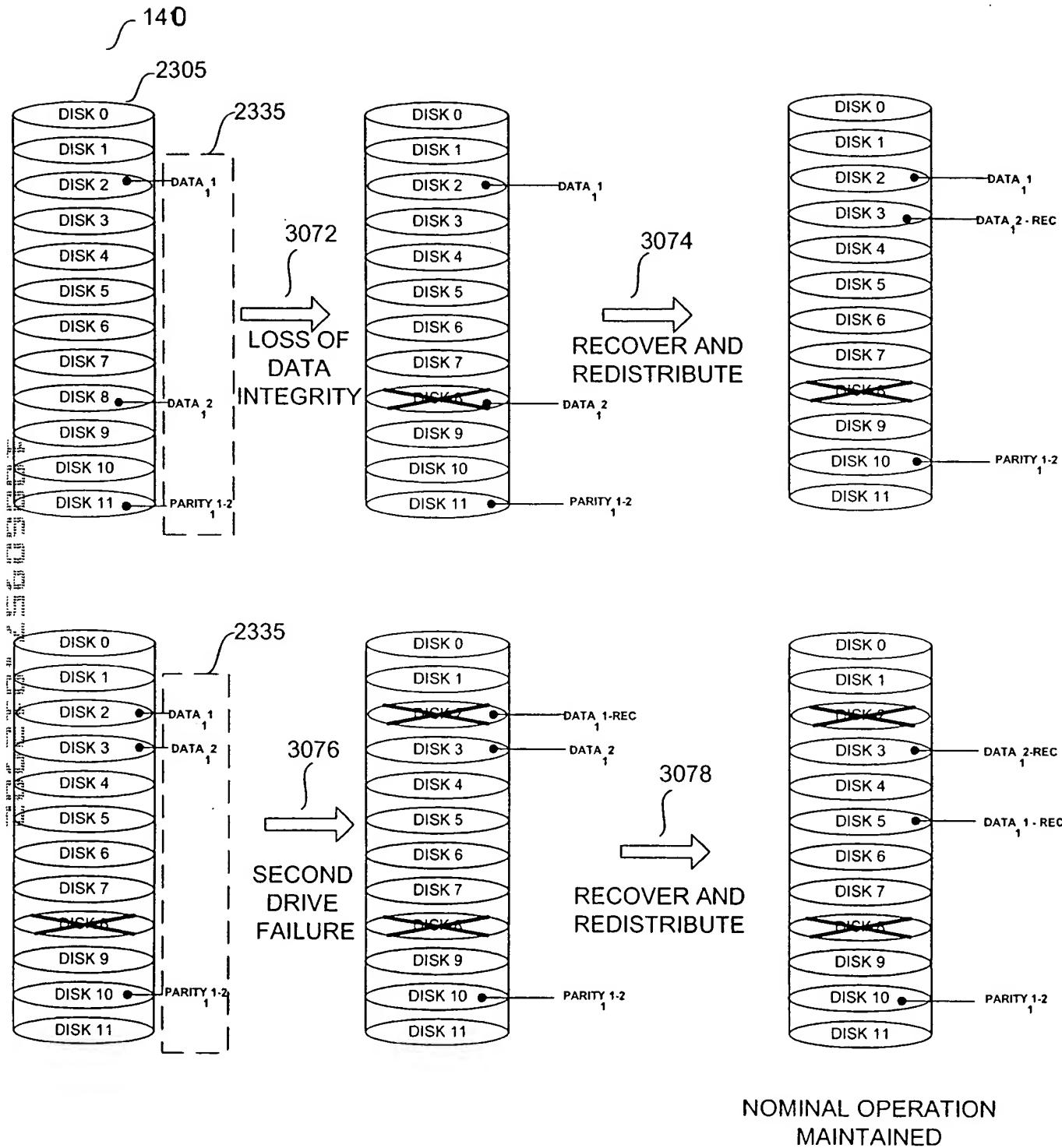
FIGURE 30

DATA RECOVERY
PROCESS
/—3172

IDENTIFY DATA OR DRIVE FAULT /
DETERMINE AFFECTED BLOCKS
/—3175

DETERMINE ASSOCIATED PARITY
GROUP
/—3177

IDENTIFY AFFECTED BLOCKS IN
G-TABLE
/—3179

DETERMINE AVAILABLE DISK
SPACE
/—3180

UPDATE G-TABLE / REBUILD DATA
/—3181

FIGURE 31

FILE #1

```
┌─────────────────────────────────────────┐
│                                         │
└─────────────────────────────────────────┘
0                                        4096
```

FILE #1 W/ PARITY -- 4-BLOCK PARITY GROUP -- EXTENT = 2   ～ 3240
5120 BYTES TOTAL / UTILIZATION = 100%

```
┌─────────────────────────────────────────┬─────────┐
│                                         │         │～3245
├───────┬───────┬───────┬───────┬─────────┤         │
│ DATA  │ DATA  │ DATA  │ DATA  │ PARITY  │         │
└───────┴───────┴───────┴───────┴─────────┘
0                                   4096
```

FILE #1 W/ PARITY -- 3-BLOCK PARITY GROUP -- EXTENT = 2   ～ 3241
8192 BYTES TOTAL / UTILIZATION = 66%

```
┌──────────────────────────┬──────┐  ┌──────────────────────────┬──────┐
│                          │      │  │                          │      │
│                        ～3247   │  │                        ～3246   │
├───────┬───────┬───────┬─┴──────┤  ├───────┬────────┬────────┬─┴──────┤
│ DATA  │ DATA  │ DATA  │ PARITY │  │ DATA  │ UNUSED │ UNUSED │ PARITY │
└───────┴───────┴───────┴────────┘  └───────┴────────┴────────┴────────┘
```

FILE #1 W/ PARITY -- 2-BLOCK PARITY GROUP -- EXTENT = 1   ～ 3242
6144 BYTES TOTAL / UTILIZATION = 100%

```
┌───────────┬──────┐ ┌───────────┬──────┐ ┌───────────┬──────┐ ┌───────────┬──────┐
│           │      │ │           │      │ │           │      │ │           │      │
├─────┬─────┼──────┤ ├─────┬─────┼──────┤ ├─────┬─────┼──────┤ ├─────┬─────┼──────┤
│DATA │DATA │PARITY│ │DATA │DATA │PARITY│ │DATA │DATA │PARITY│ │DATA │DATA │PARITY│
└─────┴─────┴──────┘ └─────┴─────┴──────┘ └─────┴─────┴──────┘ └─────┴─────┴──────┘
```

FILE #1 W/ PARITY -- 1-BLOCK PARITY GROUP -- EXTENT = 1   ～ 3243
8192 BYTES TOTAL / UTILIZATION = 100%

```
┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐ ┌──────┬──────┐
│      │      │ │      │      │ │      │      │ │      │      │ │      │      │ │      │      │ │      │      │ │      │      │
├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤ ├──────┼──────┤
│ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│ │ DATA │PARITY│
└──────┴──────┘ └──────┴──────┘ └──────┴──────┘ └──────┴──────┘ └──────┴──────┘ └──────┴──────┘ └──────┴──────┘ └──────┴──────┘
```

FILE #2

```
┌───────────┐
│           │
└───────────┘
0         1024
```

FILE #2 W/ PARITY -- 4-BLOCK PARITY GROUP -- EXTENT = 2   ～ 3250
5120 BYTES TOTAL / UTILIZATION = 25%

```
┌─────────────────────────────────────────┬─────────┐
│                                         │         │ 3245
├────────┬────────┬────────┬──────┬───────┤         │
│ UNUSED │ UNUSED │ UNUSED │ DATA │ PARITY│         │
└────────┴────────┴────────┴──────┴───────┘
```

FILE #2 W/ PARITY -- 3-BLOCK PARITY GROUP -- EXTENT = 2   ～ 3251
4096 BYTES TOTAL / UTILIZATION = 33%

```
┌──────────────────────────┬──────┐
│                          │      │
├────────┬────────┬────────┼──────┤
│ UNUSED │ UNUSED │  DATA  │PARITY│
└────────┴────────┴────────┴──────┘
```

FILE #2 W/ PARITY -- 2-BLOCK PARITY GROUP -- EXTENT = 1   ～ 3252
1536 BYTES TOTAL / UTILIZATION = 100%

```
┌───────────┬──────┐
│           │      │
├─────┬─────┼──────┤
│DATA │DATA │PARITY│
└─────┴─────┴──────┘
```

FILE #2 W/ PARITY -- 1-BLOCK PARITY GROUP -- EXTENT = 1   ～ 3253
2048 BYTES TOTAL / UTILIZATION = 100%

```
┌──────┬──────┐ ┌──────┬──────┐
│      │      │ │      │      │
├──────┼──────┤ ├──────┼──────┤
│ DATA │PARITY│ │ DATA │PARITY│
└──────┴──────┘ └──────┴──────┘
```

```
         ⌒3360

┌─────────────────────────┐ ⌒3361
│     RECEIVE FILE DATA    │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐ ⌒3363
│      ASSES FILE DATA     │
│     CHARACTERISTICS      │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐ ⌒3365
│  IDENTIFY AVAILABLE PARITY │
│    GROUP CONFIGURATIONS    │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐ ⌒3367
│    SELECT PARITY GROUP    │
│   CONFIGURATION BASED ON   │
│ UTILIZATION CHARACTERISTICS│
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐ ⌒3369
│     STORE FILE DATA IN     │
│    SELECTED PARITY GROUP   │
└─────────────────────────┘
```

FIGURE 33

FIGURE 34A

INITIAL ALLOCATION ⌐3491   DISK SPACE % ⌐3485

| DATA | DATA | DATA | DATA | PARITY | 4 block parity ⌐3480   10000 groups   36%

| DATA | DATA | DATA | PARITY | 3 block parity ⌐3481   10000 groups   28%

| DATA | DATA | PARITY | 2 block parity ⌐3482   10000 groups   22%

| DATA | PARITY | 1 block parity ⌐3483   10000 groups   14%

DISK USAGE ⌐3487

FIGURE 34B

3492 FREE	OCCUPIED ⌐3490 TOTAL	DISK SPACE %

3480 4 block parity   2500 groups   7500 groups   10000 groups   36%

3481 3 block parity   7500 groups   2500 groups   10000 groups   28%

3482 2 block parity   3500 groups   6500 groups   10000 groups   22%

3483 1 block parity   500 groups   9500 groups   10000 groups   14%

REDISTRIBUTION ⌐3494

FIGURE 34C

3492 FREE	OCCUPIED ⌐3490 TOTAL	DISK SPACE %

3480 4 block parity   2500 groups   7500 groups   10000 groups   36%

3481 3 block parity   -5000 groups of 3 block parity   2500 groups   2500 groups   5000 groups   14%

3482 2 block parity   3500 groups   6500 groups   10000 groups   22%

3483 1 block parity   +10000 groups of 1 block parity   10500 groups   9500 groups   20000 groups   28%

REDISTRIBUTION

PARITY GROUP REDISTRIBUTION PROCESSES $\int$ 3500

FIGURE 35A   PARITY GROUP DISSOLUTION $\int$ 3510
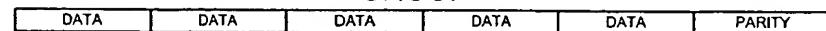
5-BLOCK PARITY
GROUP $\int$ 3515

DATA	DATA	DATA	DATA	DATA	PARITY

1-BLOCK PARITY $\int$ 3520
GROUP

DATA	PARITY

3-BLOCK PARITY $\int$ 3525
GROUP

DATA	DATA	DATA	PARITY

OR

2-BLOCK PARITY $\int$ 3530
GROUP

DATA	DATA	PARITY

2-BLOCK PARITY $\int$ 3530
GROUP

DATA	DATA	PARITY

OR

1-BLOCK PARITY $\int$ 3520
GROUP

DATA	PARITY

1-BLOCK PARITY $\int$ 3520
GROUP

DATA	PARITY

1-BLOCK PARITY $\int$ 3520
GROUP

DATA	PARITY
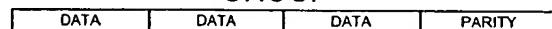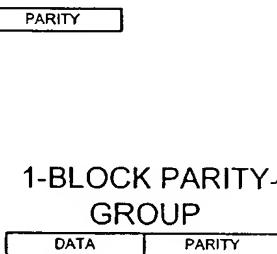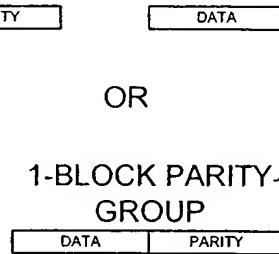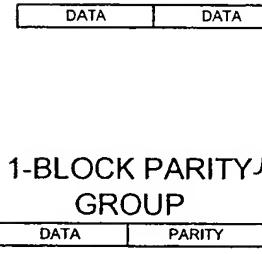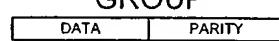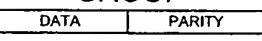
FIGURE 35B $\int$ 3535

PARITY GROUP CONSOLIDATION

3-BLOCK PARITY GROUP $\int$ 3525

DATA	DATA	DATA	PARITY

2-BLOCK PARITY
GROUPS $\int$ 3530

DATA	DATA	PARITY

DATA	DATA	PARITY

1-BLOCK PARITY $\int$ 3520
GROUP

DATA	PARITY

OR

5-BLOCK PARITY GROUP $\int$ 3515

DATA	DATA	DATA	DATA	DATA	PARITY

_–3600

_–3602

MONITOR PARITY GROUP
AVAILABILITY

_–3604

PARITY GROUP
DEPLETION
OBSERVED
?

NO

YES

_–3606

ASSESS PARITY GROUP
STATISTICS

_–3608

REDISTRIBUTE PARITY GROUPS
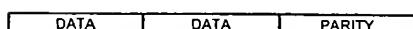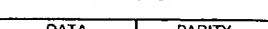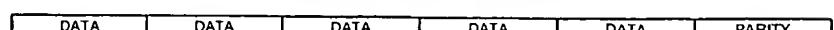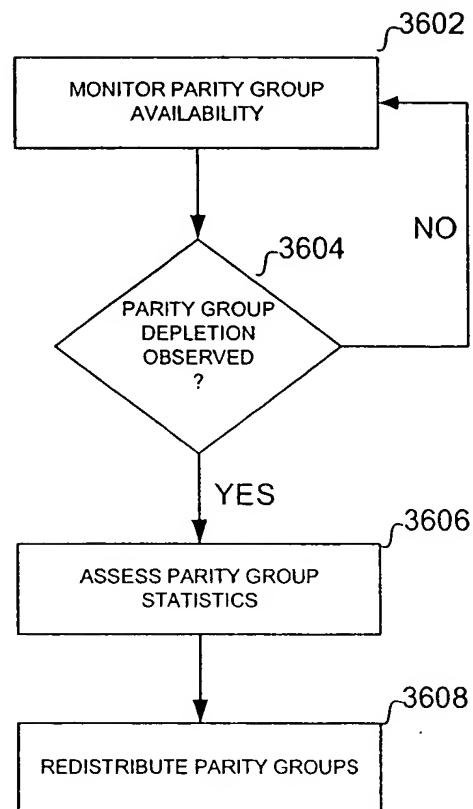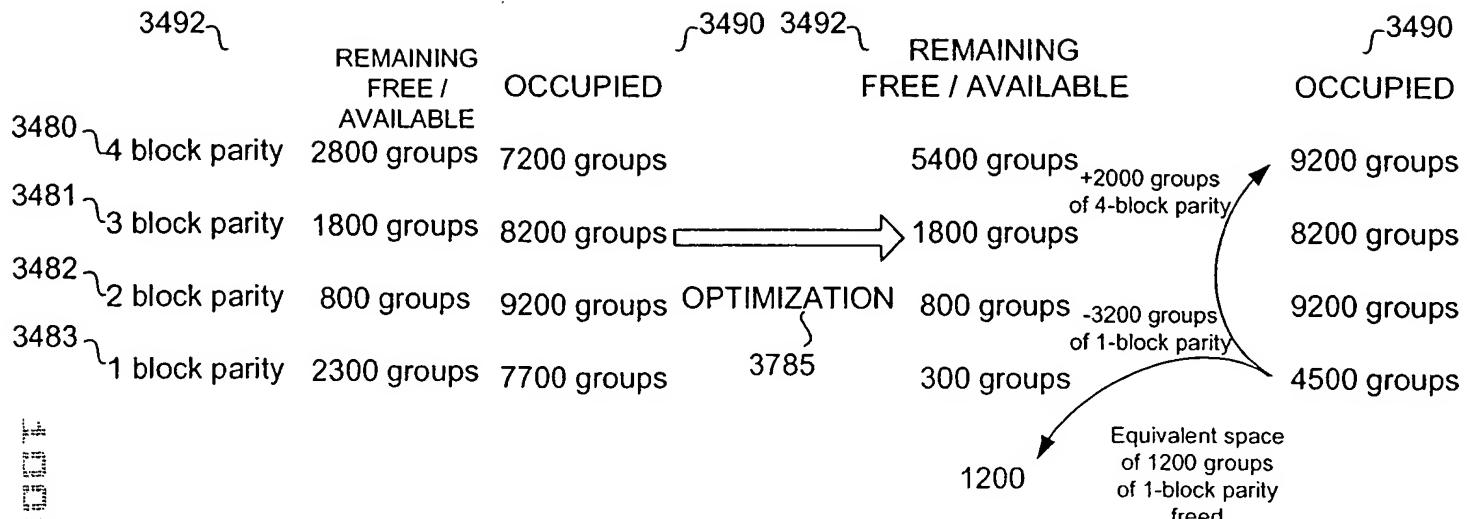
FIGURE 36

FIGURE 37

FIGURE 38

FIG. 39

SERVER
INTERFACE
4005

FILE SYSTEM
4010

NETWORK DRIVERS
AND STACKS

INTER-NODE COMM	BACKUP	ADMIN INTERFACE	PROTOCOL INTERPRETER	PROTOCOL INTERPRETER

SERVER MANAGER   4008

REQUEST PROCESSING   4012

DATA AND METADATA MGMT   4014

STORAGE MGMT   4016

NETWORK DRIVER

OBJECT POSITIONER   4018

OBJECT POSITIONING PLAN   4025

STATISTICS

F3 LOAD BALANCING DATA

LOAD BALANCING DATA   4020

F1 LOAD BALANCING DATA	F2 LOAD	F4 LOAD	F5 LOAD
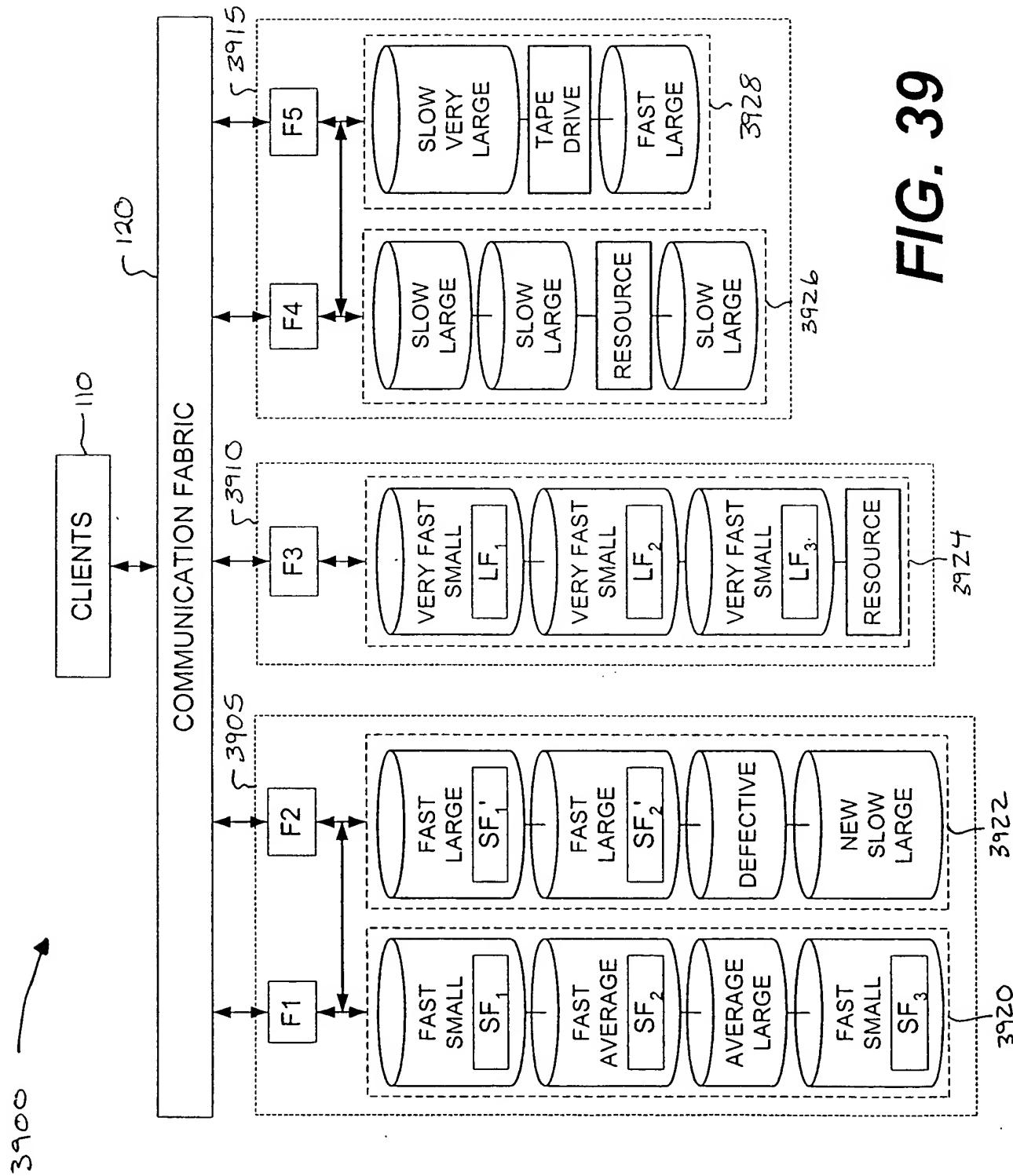
4000

FIG. 40
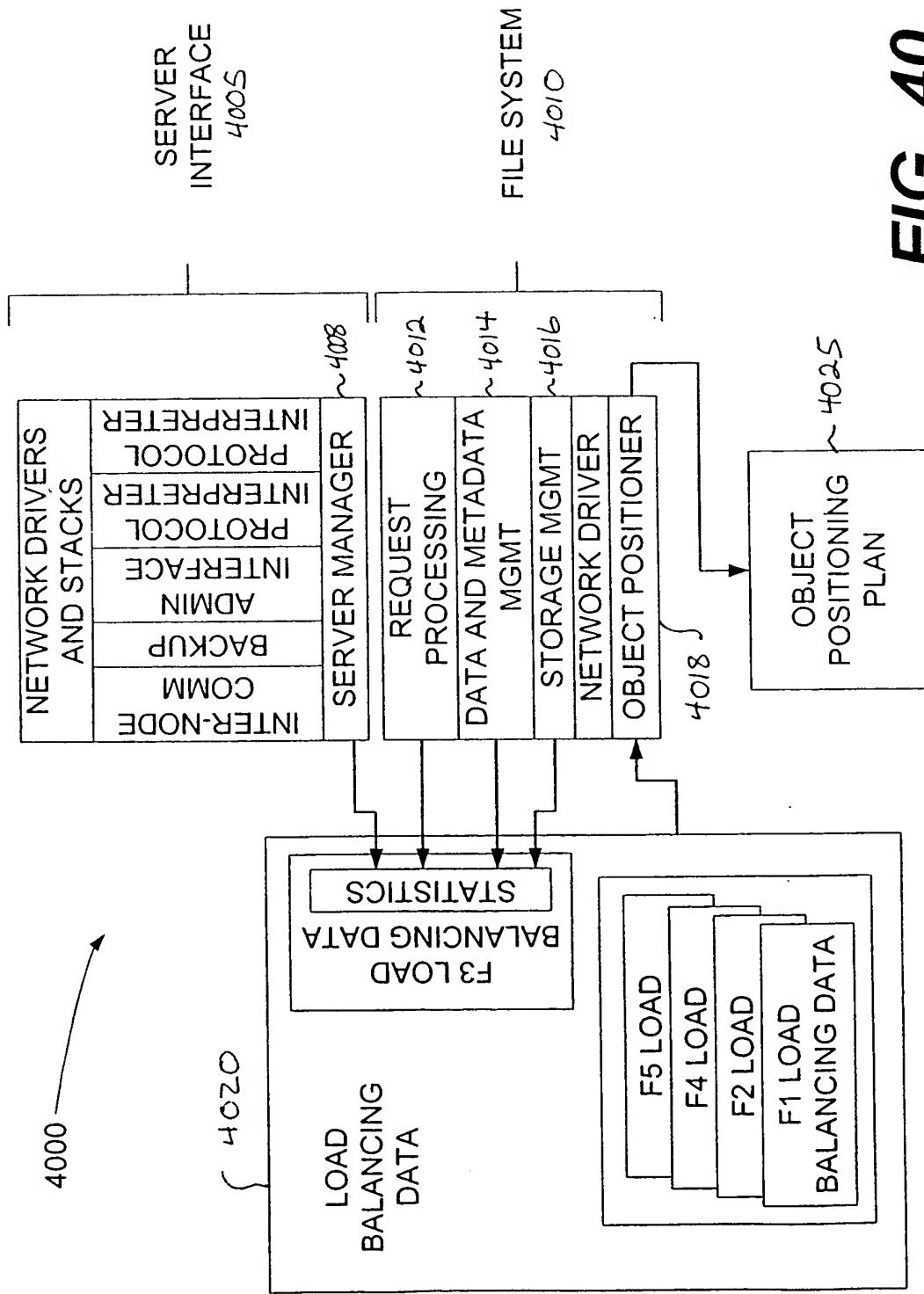
F3 OBJECT
POSITIONING PLAN

-Push LF to F4-F5 Cluster

-Issue File Handle For LF = Stale

-If Requested,
  -Send acceptance for copy
    of SF to F1
  -Create copy of SF
  -Send file handle of SF to F1

FIG. 41

4025

FIGURE 42

Clients — 110

Network Interface — 4214

Network Interface

4201

Network Bus

CPU(s) — 4205

File System Metadata RAM Non Volatile — 4216

DATA ENGINE — 4210

Data/Parity Cache RAM A Non Volatile — 4218

Data/Parity Cache RAM B Non Volatile — 4220

4202

Storage Bus

Storage Interface — 4212

Storage Interface

Disks — 140,141

130
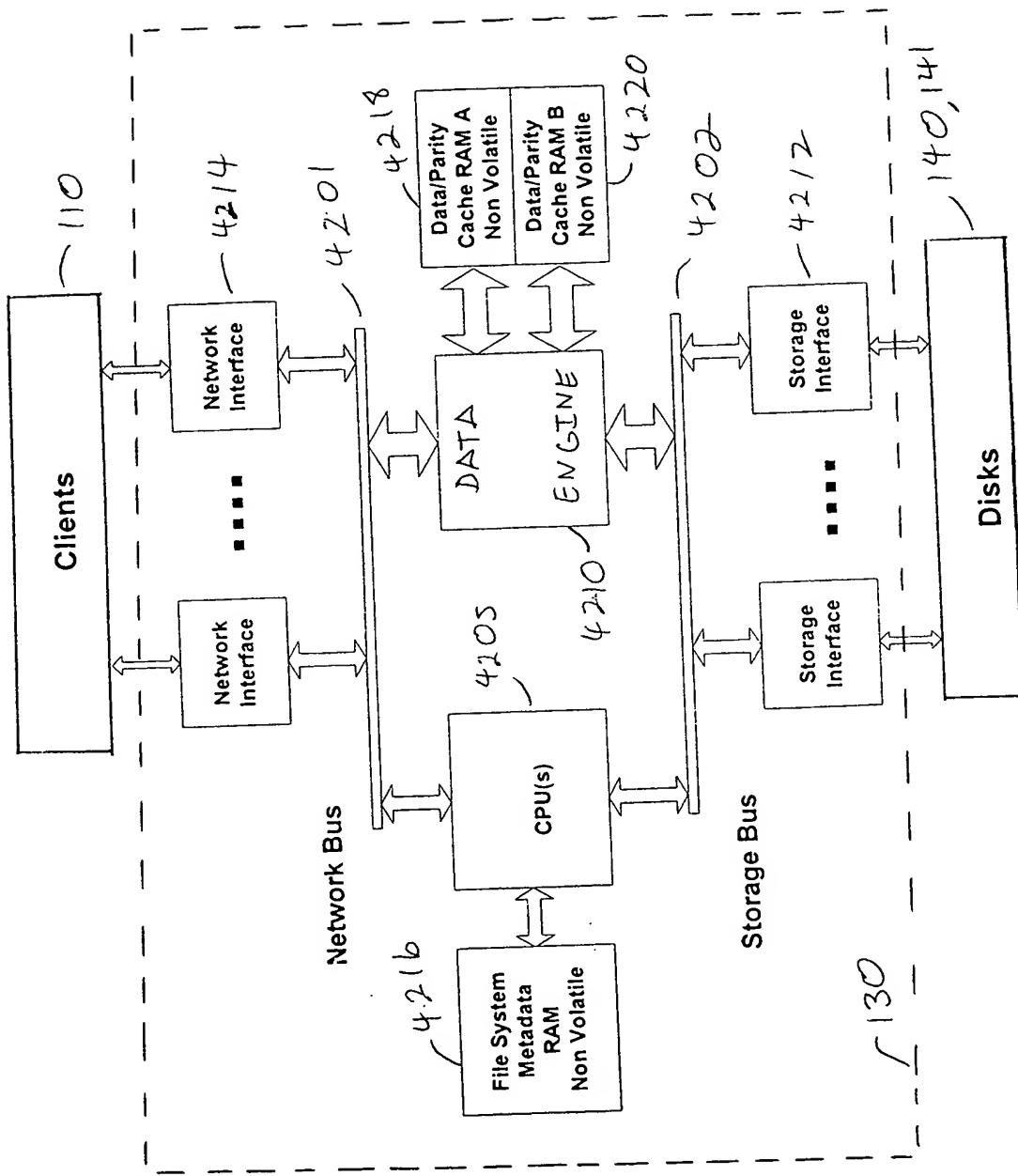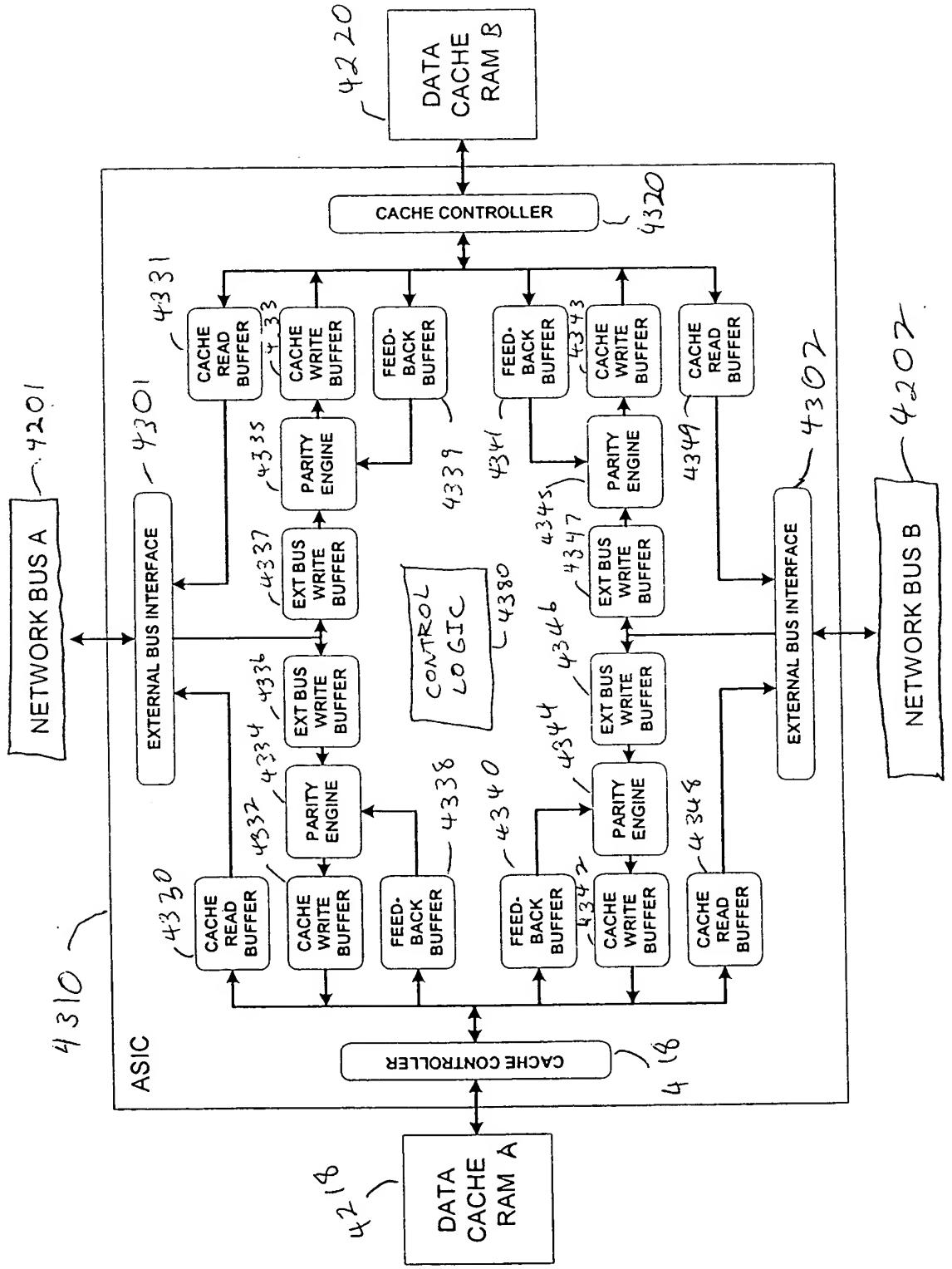
FIGURE 43

| PCI map | Block Size | Opcode | Spare | Parity Index | Spare | RAM Adr |

63 - - - 62,61 - - - - - 59,58 - - - - - - 56,55 - - - - - 51,50 - - - - 35,34,32, 31 - - - - - - - - - 0

FIGURE 44

4400